| | |
|---|---|
| **CS 880: Advanced Complexity Theory** | 2/4/2008 |

# Lecture 6: Influence and Noise Sensitivity

Instructor: Dieter van Melkebeek Scribe: David Malec

In the past few lectures we studied testing closeness of a Boolean function to classes such as linear functions and dictator functions. The basis was a simple connection between the Fourier spectrum of a given function $f$ and its distance to the function class LINEAR. In this lecture we look at two more natural properties of Boolean functions, influence and noise sensitivity, and develop simple expressions for them based on the Fourier spectrum.

Recall that for a given Boolean function $f : \{-1, 1\} \to \{-1, 1\}$, we define the influence of the $i^{th}$ variable as

$$I_i = \Pr_x \left[ f(x^{(i)}) \neq f(x) \right],$$

the total influence as

$$I = \sum_{i=1}^{n} I_i,$$

and the noise sensitivity of $f$ at $\epsilon$ as

$$\mathrm{NS}_\epsilon = \Pr_{\substack{x \\ y \sim_\epsilon x}} \left[ f(x) \neq f(y) \right],$$

where we use the notation $y \sim_\epsilon x$ to denote that $y$ is formed from $x$ by flipping each bit of $x$ independently with probability $\epsilon$.

## 1 Influence

If we want to express the influence of a function, we could follow the same method used in previous lectures. Namely, for a Boolean function $f$, we have $f(x) = f(y)$ if and only if $f(x)f(y) = 1$, and so we can write

$$\begin{aligned} \mathop{\mathrm{E}}_x \left[ f(x^{(i)}) f(x) \right] &= \Pr_x \left[ f(x^{(i)}) = f(x) \right] - \Pr_x \left[ f(x^{(i)}) \neq f(x) \right] \\ &= 1 - 2 \cdot \Pr_x \left[ f(x^{(i)}) \neq f(x) \right], \end{aligned}$$

and so by the definition of $I_i$,

$$\mathop{\mathrm{E}}_x \left[ f(x^{(i)}) f(x) \right] = 1 - 2 I_i. \tag{1}$$

We could then derive an expression for $I_i$ by substituting the Fourier expansion of $f$ into the left-hand side of (1); we do not follow that approach here, but leave it as an exercise for the reader.

## 1.1 The Difference Operator

Instead, we shall build our analysis around the difference operator $D_i$, defined by

$$(D_i f)(x) = \frac{f(x^{(i=1)}) - f(x^{(i=-1)})}{2},$$

where we use $x^{(i=1)}$ ($x^{(i=-1)}$) to indicate $x$ with the $i^{th}$ bit set to 1 ($-1$). The difference function can be thought of as providing a discrete analogue to the partial derivative of $f$ with respect to $x_i$. The factor of 2 in the denominator can be interpreted as the difference $x_i^{(i=1)} - x_i^{(i=-1)}$.

## 1.2 Using $D_i$ to Express the Influence of a Function

Considering the definition of $D_i f$, we note the following. Given any $x$, if changing the $i^{th}$ bit of $x$ does not change the value of $f(x)$, then $(D_i f)(x) = 0$; on the other hand, if changing the $i^{th}$ bit does affect the function value, then we have that $(D_i f)(x) = \pm 1$. So we can express the influence of $f$ in terms of the absolute value of $D_i f$; specifically, we have

$$I_i = \mathop{\mathrm{E}}_x \left[ |(D_i f)(x)| \right]. \tag{2}$$

Working from Equation (2), we may arrive at the following theorem.

**Theorem 1.** *Given a Boolean function $f$, $I(f) = \sum_{S \subseteq [n]} |S| \left( \widehat{f}(S) \right)^2$.*

*Proof.* We work with the Fourier expansion of $D_i f$; before proceeding with our analysis, however, we make a couple of observations that simplify our task. First, we note that if $i \notin S$, then flipping the $i^{th}$ bit of $x$ does not change the value of $\chi_S(x)$; formally, we have that

$$\forall i \notin S, \chi_S(x^{(i=-1)}) = \chi_S(x^{(i=1)}). \tag{3}$$

Second, given a set $S$ and any $i \in S$, we can separate out the factor of $\chi_S$ corresponding to $i$; this follows from the definition

$$\chi_S(x) = \prod_{j \in S} x_j = \left( \prod_{\substack{j \in S \\ j \neq i}} x_j \right) x_i.$$

The importance of this is that if we know $i \in S$, then we can say that

$$\begin{aligned} \chi_S(x^{(i=1)}) - \chi_S(x^{(i=-1)}) &= 1 \cdot \chi_{S \setminus \{i\}}(x) - (-1) \cdot \chi_{S \setminus \{i\}}(x) \\ &= 2\chi_{S \setminus \{i\}}(x) \end{aligned} \tag{4}$$

Using these two observations, we proceed as follows.

$$\begin{aligned} (D_i f)(x) &= \sum_{S \subseteq [n]} \widehat{f}(S) \frac{\chi_S(x^{(i=1)}) - \chi_S(x^{(i=-1)})}{2} \\ &= \sum_{\substack{S \subseteq [n] \\ S \ni i}} \widehat{f}(S) \frac{\chi_S(x^{(x=1)}) - \chi_S(x^{(x=-1)})}{2} && (5) \\ &= \sum_{S \ni i} \widehat{f}(S) \chi_{S \setminus \{i\}}(x), && (6) \end{aligned}$$

where (5) follows from (3), and (6) follows from (4). Since $\mathrm{Ran}(D_i f) = \{-1, 0, 1\}$, we have that $|D_i f| = (D_i f)^2$, and so

$$\begin{aligned}
\mathop{\mathrm{E}}_x [|(D_i f)(x)|] &= \mathop{\mathrm{E}}_x \left[ (D_i f)^2 (x) \right] \\
&= \sum_{S \subseteq [n]} \left( \widehat{D_i f} \right)^2 (S) && (7) \\
&= \sum_{S \ni i} \left( \widehat{f}(S) \right)^2, && (8)
\end{aligned}$$

where (7) follows from Parseval's Equality, and (8) follows from (6). Hence, we get that

$$\begin{aligned}
I(f) &= \sum_{i=1}^{n} \sum_{S \ni i} \left( \widehat{f}(S) \right)^2 \\
&= \sum_S |S| \left( \widehat{f}(S) \right)^2 .
\end{aligned}$$

□

# 2 Applications of Influence

We now present two applications where the bound from Theorem 1 proves useful.

## 2.1 Influence vs. Variance

The first application we consider deals with bounding the variance of a function by the sensitivity. Working from the definition of $\sigma^2(f)$, we arrive at the following proposition.

**Proposition 1.** *Given a Boolean function $f$, $\sigma^2(f) \leq I$.*

*Proof.*

$$\begin{aligned}
\sigma^2(f) &= \mathop{\mathrm{E}}_x \left[ \left( f(x) - \mathop{\mathrm{E}}_x [f] \right)^2 \right] \\
&= \mathop{\mathrm{E}}_x \left[ (f(x))^2 \right] - \left( \mathop{\mathrm{E}}_x [f(x)] \right)^2 \\
&= \sum_{S \subseteq [n]} \left( \widehat{f}(S) \right)^2 - \left( \widehat{f}(\emptyset) \right)^2 && \text{(by Parseval's)} \\
&= \sum_{S \neq \emptyset} \left( \widehat{f}(S) \right)^2 \\
&\leq \sum_S |S| \left( \widehat{f}(S) \right)^2 \\
&= I. && \text{(by Theorem 1)}
\end{aligned}$$

□

Now, consider the expression $\mathop{\mathrm{E}}_x\left[f^2(x)\right] - \left(\mathop{\mathrm{E}}_x[f(x)]\right)^2$, which we used in the proof of Proposition 1. Since we have that $f$ is a Boolean function, we can see that $f^2(x)$ must always have a value of 1; furthermore, if we let $p$ be the probability that $f(x) = -1$ for uniform $x$, we can compute the variance of $f$ in terms of $p$ as

$$\begin{aligned}\sigma^2(f) &= \mathop{\mathrm{E}}_x\left[f^2(x)\right] - \left(\mathop{\mathrm{E}}_x[f]\right)^2 \\ &= 1 - (p \cdot (-1) + (1-p) \cdot 1)^2 \\ &= 1 - (1-2p)^2 \\ &= 4p(1-p).\end{aligned}$$

Combining this with Proposition 1, we can see that

$$I(f) \geq 4p(1-p). \tag{9}$$

For a balanced function, we know that we have $p = 1/2$; making this substitution in (9), we immediately arrive at the following corollary.

**Corollary 1.** *For a balanced Boolean function $f$, $I(f) \geq 1$.*

Note that most of the provided proof for Proposition 1 made no use of the fact that $f$ was Boolean; we only used it to switch between $|D_i f|$ and $(D_i f)^2$. Proposition 1 can be viewed as a special case of Poincaré's Inequality, which has the form

$$\mathop{\mathrm{E}}_x\left[\|\overrightarrow{\nabla} f\|^2\right] \geq c \cdot \sigma^2(f)$$

This inequality holds in certain spaces, called Sobolev spaces; for such spaces, the constant required is called the Poincaré constant for the space. In the case of the Boolean cube, the constant is 1.

## 2.2 Monotone Functions

Our second application is to monotone functions. If $f : \{-1,1\}^n \to \{-1,1\}$ is monotone, then for any $x \in \{-1,1\}^n$ we know that

$$f(x^{(i=-1)}) \leq f(x^{(i=1)}),$$

and so

$$f(x^{(i=1)}) - f(x^{(i=-1)}) \geq 0$$

from which we may conclude that

$$(D_i f)(x) \geq 0.$$

If we revisit our original analysis of $I_i$, we can see that for monotone $f$ we need not take an absolute value within the expected value, and so by (6) get

$$I_i = \mathop{\mathrm{E}}_x[|D_i f|] = \mathop{\mathrm{E}}_x[D_i f] = \widehat{D_i f}(\emptyset) = \widehat{f}(\{i\}),$$

from which we may conclude that

$$I = \sum_{i=1}^{n} \widehat{f}(\{i\}). \tag{10}$$

If we consider the right-hand side of (10), we see that we can view it as the inner product of a vector of $f$-values and the all 1's vector of dimension $n$. We can apply the Cauchy-Schwartz Inequality to derive an upper bound for $I(f)$ as follows.

$$\begin{aligned} I &\le \sqrt{n}\sqrt{\sum_{i=1}^{n}\left(\widehat{f}(\{i\})\right)^2} \\ &\le \sqrt{n}\sqrt{\sum_{S}\left(\widehat{f}(S)\right)^2} \\ &= \sqrt{n}. \end{aligned}$$

Thus, a monotone function has sensitivity no more than $\sqrt{n}$. In fact, we can say more; we can derive a tighter bound, and even determine when the bound becomes tight.

**Proposition 2.** *For monotone Boolean functions $f$, $I(f)$ is maximized for* Majority*.*

*Proof.* First, we derive the following upper bound for monotone Boolean functions $f$.

$$\begin{aligned} I &= \sum_{i=1}^{n} \widehat{f}(\{i\}) \\ &= \sum_{i=1}^{n} \mathop{\mathrm{E}}_{x}\left[f \cdot x_i\right] && (11) \\ &= \mathop{\mathrm{E}}_{x}\left[f \cdot \sum_{i=1}^{n} x_i\right] && (12) \\ &\le \mathop{\mathrm{E}}_{x}\left[|\sum_{i=1}^{n} x_i|\right]. && (13) \end{aligned}$$

In the above, (11) follows from the definition of the Fourier coefficients, and (12) follows from linearity of expectation. Note that (13) comes from the fact that since $f$ maps to the range $\{-1, 1\}$, we have that $|f| = 1$; furthermore, this inequality becomes tight iff

$$f \cdot \sum_{i=i}^{n} x_i = |\sum_{i=1}^{n} x_i|.$$

We can make use of the sign function to rewrite this as

$$f \cdot \sum_{i=i}^{n} x_i = \mathrm{sgn}\left(\sum_{i=1}^{n} x_i\right)\sum_{i=1}^{n} x_i,$$

and so see that (13) becomes tight iff

$$f = \operatorname{sgn}\left(\sum_{i=1}^{n} x_i\right), \text{ whenever } \sum_{i=1}^{n} x_i \neq 0. \tag{14}$$

Since each $x_i = \pm 1$, the sign of $\sum_{i=1}^{n} x_i$ is simply the sign shared by the majority of the $x_i$'s, and so (14) holds for $f = \text{Majority}$, regardless of how we define the behavior of Majority when equal numbers of $x_i$'s take on the values 1 and $-1$. $\square$

Note that if we apply Sterling's Formula, we can then determine the factor which was lost in our earlier analysis, finding that

$$I(\text{Majority}) \sim \sqrt{\frac{2}{\pi}}\sqrt{n}.$$

# 3 Noise Sensitivity

Here, we work with the equation

$$\mathop{\mathrm{E}}_{\substack{x \\ y\sim_\epsilon x}} [f(x)f(y)] = 1 - 2\,\mathrm{NS}_\epsilon(f) \tag{15}$$

In order to simplify our work, we introduce the noise operator $T_\alpha$, defined by

$$(T_\alpha f)(x) = \mathop{\mathrm{E}}_{y\sim_\epsilon x} [f(y)], \text{ where } \epsilon = \frac{1-\alpha}{2}.$$

One important aspect of the noise operator is that both its domain and range are equal to the class of functions $\{f : \{-1,1\}^n \to \mathbb{R}\}$; applying $T_\alpha$ to a Boolean function does not necessarily yield another Boolean function. Note also the similarity to the operator used in the dictator test from Lecture 5. We get the same relationship seen in that analysis between Fourier coefficients, namely

$$\widehat{T_\alpha f}(S) = \alpha^{|S|}\widehat{f}(S). \tag{16}$$

Using the operator $T_\alpha$, we can now see that

$$\begin{aligned}
1 - 2\,\mathrm{NS}_\epsilon &= \mathop{\mathrm{E}}_{x}[f(x)(T_\alpha f)(x)] \\
&= \sum_{S\subseteq[n]} \widehat{f}(S)\widehat{T_\alpha f}(S) && \text{(by Plancherel's)} \\
&= \sum_{S\subseteq[n]} \widehat{f}(S)\alpha^{|S|}\widehat{f}(S) && \text{(by Equation (16))} \\
&= \sum_{S\subseteq[n]} \alpha^{|S|}\left(\widehat{f}(S)\right)^2,
\end{aligned}$$

and so we get that

$$\mathrm{NS}_\epsilon = \frac{1}{2} - \frac{1}{2}\sum_{S}(1-2\epsilon)^{|S|}\left(\widehat{f}(S)\right)^2.$$

From this analysis, we can derive a bound for the noise sensitivity of $f$ in terms of the total influence of $f$.

**Proposition 3.** $\mathrm{NS}_\epsilon(f) \leq \epsilon \cdot I(f)$.

*Proof.* First, we can bound the value of $(1-2\epsilon)^{|S|}$, by expanding out the product and truncating it to the first two terms, to get

$$(1-2\epsilon)^{|S|} \geq 1 - 2\epsilon|S|,$$

from which we can see that

$$\begin{aligned}
\mathrm{NS}_\epsilon(f) &= \frac{1}{2} - \frac{1}{2}\sum_S (1-2\epsilon)^{|S|}\left(\widehat{f}(S)\right)^2 \\
&\leq \frac{1}{2} - \frac{1}{2}\sum_S (1-2\epsilon|S|)\left(\widehat{f}(S)\right)^2 \\
&= \epsilon \cdot \sum_S |S|\left(\widehat{f}(S)\right)^2 \\
&= \epsilon \cdot I(f).
\end{aligned}$$

□

## 4 Hypercontractivity of the Noise Operator $T_\alpha$

Here, we briefly discuss an important property of the operator $T_\alpha$, which is critical to proving:

- for every balanced $f$, $I_i(f) \geq \Omega(\frac{\log n}{n})$ for some $i$;
- for every Boolean function $f$, $\forall \epsilon > 0$, $f$ is $\epsilon$-close to some $r$-junta, where $r = 2^{O(I/\epsilon)}$.

The property referred to is hypercontractivity. Say we have an operator $T$ which acts on functions $f : \{-1,1\} \to \mathbb{R}$. Contractivity of $T$ means that applying the operator to a function $f$ does not increase the function's norm. Recall that a norm on a space $V$ is a function $\|v\|$ from $V$ to $\mathbb{R}^+$ such that the following hold for all $u, v \in V$:

1. $\|v\| \geq 0$, and $\|v\| = 0$ iff $v = 0$;
2. $\|\alpha \cdot v\| = |\alpha| \cdot \|v\|$ for any scalar $\alpha$;
3. $\|u + v\| \leq \|u\| + \|v\|$.

The latter property is referred to as the triangle inequality.

The norms we consider are of the form

$$\|f\|_p = \left(\mathrm{E}\left[|f|^p\right]\right)^{1/p}, \text{ where } 1 \leq p \leq \infty.$$

Properties 1 and 2 are easily verified and hold for each $p > 0$. The triangle inequality takes more work and only holds for $p \geq 1$.

We can see how this family of norms compare to each other by looking at some examples of the unit circles they produce in the case where $n = 1$ (see Figure 1). Note that whenever $|f|$ is constant, say $|f| \equiv a$, then $\|f\|_p$ is independent of $p$ and equal to $a$. Thus, in Figure 1, all unit

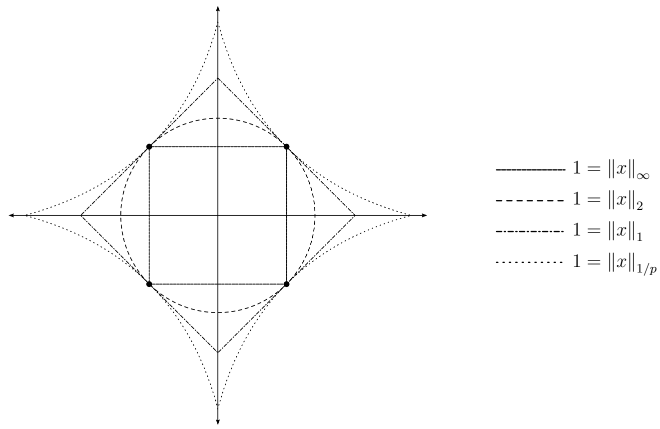


Figure 1: Unit Circles Under $p$-Norms

circles share the 4 points $(\pm 1, \pm 1)$. When we let $p$ decrease from $\infty$ over 2 downto 1, the unit circle morphs from a square over an actual circle to a diamond, with the 4 points $(\pm 1, \pm 1)$ as fixed points. When we decrease $p$ below 1, the "unit circle" becomes non-convex, which is exactly because $\|\cdot\|_p$ no longer satisfies the triangle inequality.

From Figure 1, we can see that for any $f$, $\|f\|_p \leq \|f\|_q$ whenever $p \leq q$. We leave it as an exercise to prove this formally.

Under these norms, the operator $T_\alpha$ has the property of hypercontractivity, formalized by the following.

**Theorem 2.** $\forall 1 \leq p \leq q \leq \infty, \forall |\alpha| \leq \sqrt{\frac{p-1}{q-1}}$, *it is the case that*

$$\|T_\alpha f\|_q \leq \|f\|_p.$$

To say that $T_\alpha$ has the property of contractivity means that $\|T\alpha f\|_p \leq \|f\|_p$; that is, applying $T_\alpha$ to a function $f$ never increases the $p$-norm of $f$. Theorem 2 states something stronger: not only do we maintain the relationship that the norm of $T_\alpha f$ is at most the norm of $f$, but this statement even holds when we change the norm we use for $T_\alpha f$ to one which generally has larger values, namely the $q$-norm instead of the $p$-norm for $q > p$. Hence the term "hypercontractivity."